

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: TUCKER et al.
- (ii) TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
- (iii) NUMBER OF SEQUENCES: 21
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: PENNIE & EDMONDS LLP
(B) STREET: 1155 Avenue of the Americas
(C) CITY: New York
(D) STATE: New York
(E) COUNTRY: USA
(F) ZIP: 10036-2711
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/968,685
(B) FILING DATE: November 12, 1997
(C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Baldwin, Geraldine F.
(B) REGISTRATION NUMBER: 31,232
(C) REFERENCE/DOCKET NUMBER: 7969-060
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212) 790-9090
(B) TELEFAX: (212) 869-8864
(C) TELEX: 66141 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
- | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Gly | Ile | Ser | Glu | Ala | Asp | Gly | Gly | Lys | Gly | Gly | Ala | Asn | Ala | Arg |
| 1 | | | | | 5 | | | | 10 | | | | | 15 | |

Gly Asp Lys Ser Ile Ala Ile Gly Asp Ile Ala Gln Ala Leu Gly Ser
 20 25 30

Gln Ser Ile Ala Ile Gly Asp Asn Lys Ile Val
 35 40

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Gly Thr Val Leu Gly Gly Lys Lys
 1 5

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "probe"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGNACNGTNC TNGGNGGNAA RAAR

24

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 72 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..72

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAA GCG GAC GGG GGG AAA GGC GGA GCC AAT GCG CGC GGT GAT AAA TCC

48

Glu Ala Asp Gly Gly Lys Gly Gly Ala Asn Ala Arg Gly Asp Lys Ser
 1 5 10 15

72

ATT GCT ATT GGT GAC ATT GCG CAA
 Ile Ala Ile Gly Asp Ile Ala Gln
 20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Glu Ala Asp Gly Gly Lys Gly Gly Ala Asn Ala Arg Gly Asp Lys Ser
 1 5 10 15
 Ile Ala Ile Gly Asp Ile Ala Gln
 20

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "probe"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

YTTYTTNCCN CCNAGNACNG TNCC

24

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "probe"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGNACNGTNT TRGGNGGNAA RAAR

24

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = ""probe""
- (v) FRAGMENT TYPE: N-terminal
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

YTTYTTNCCN CCYAANACNG TNCC

24

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9542 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTATTGACTT AAATCACCAT ATGGTTATAA TTTAGCATAA TGGTAGGCTT TTTGTAAAAA 60
 TCACATCGCA ATATTGTTTT ACTGTTACTA CCATGCTTGA ATGACGATCC AAATCACCAG 120
 ATTCATTCAA GTGATGTGTT TGTATACGCA CCATTTACCC TAATTATTTC AATCAAATGC 180
 CTATGTCAGC ATGTATCATT TTTTAAGGTA AACCACCATG AATCACATCT ATAAAGTCAT 240
 CTTTAACAAA GCCACAGGCA CATTTATGGC CGTGGCGGAA TATGCCAAAT CCCACAGCAC 300
 GGGGGGGGGT AGCTGTGCTA CAGGGCAAGT TGGCAGTGTA CGCACTCTGA GCTTTGCCCG 360
 TATTGCCGCG CTCGCTGTCC TCGTGATCGG TGCGACGCTC AATGGCAGTG CTTATGCAGG 420
 TATCGGAATT AGTGAAGCAG ACGGGGGAAA AGGCGGAGCC AATGCGCGCG GTGATAAATC 480
 CATTGCTATT GGTGATATTG CTCAGGCACT TGGCTCTCAA TCTATTGCTA TCGGTGACAA 540
 CAAAATAGTT CATAATTCAA ATAATAATGC TAATATAGGT GCCAAAGCCT CAGGTAATGA 600
 GTCCATCGCC ATCGGTGGTG ATGTATTGGC TTCTGGTCAT GCCTCGATTG CCATCGGTAG 660
 TGATGACTTA TATTTGAAAA AGGAAACGGT ACAGCAAATC TCAGAGCTTC TACCTATTAT 720
 TCGCGGACAG AAAGCATTAA ACGATATATA CCAACTAGCT GACACTAATC TTCAAAAATA 780
 TAGACGCACA CACGCACAGG GACACGCCAG TACTGCAGTG GGAGCCATGT CATATGCAAA 840
 GGGTCATTTT TCCAACGCCT TTGGTACACG GGCAACAGCT GAAGGTACCT ATTCCTTGGC 900
 AGTGGGTCTT ACCGCCACAG CCAAAGCAGC ATCTTCAATC GCTGTTGGTT CTAATGCACA 960

AGCTATCGGG TTTGCAGCGA CAGCCGTTGG TGGAAGTACT CAAGTTAATT TGAATCGAGG 1020
 TATTGCCCTA GGTTTTGGTT CTCAGGTCCT TCAGAAGGAT AATGATGTAA ATGCAGCAAA 1080
 TGTACGGGCC TATGCACCAG ATGATAACCA GCCAATAGAC AACCGGTATA AAGCCACCTT 1140
 CAAGAATGGT GCTACGGATG TATTTTCCAT TGGTAATAGT AATGGGAATG ACAGTATCAG 1200
 GCGTAAAATC ATCAATGTCG GTGCAGGTTT TCGGATACC GATGCGGTCA ATGTGGCACA 1260
 GCTTAAAGAG GCGGTGAGGC TGGCTAATCG TCAAATTACT TTTAAGGGTG ATGATAGCAA 1320
 TAATAGAGTA GAAAAAGGTT TGGGCAAGAC TTAACTATC ACAGGTGGTG CACAGACCAG 1380
 CGCATTAACC GATCATAACA TCGGTGTGGT ACAAATGGC GATGGTCTGA AAGTTCAACT 1440
 TGCTGAAACT TTAACCAGCC TTAAATGGT TACCACTGAA AACCTAACCG CCAACGAGAA 1500
 AGTTACCGTA GGCAAAACCC GCCTTACCAC AGATAAAATT GGTTTTACCA ATGATATGAA 1560
 TGGCATTGAT GAAAGCAAAC CTTATCTTGA TAAAGACACT GGCATTCATG CAGGTGGTCA 1620
 AAAGATTACC AAAGTTACTG CTGGTGTAGT AGATGACGAT GCGGCAACTT ATGGACAGCT 1680
 TAAAAAAGTT AACCAAACCG CTGAAAGTGC TCTACAAACC TTTACCGTTA AAAAGGTAGA 1740
 TAAAAATGGT AATGATGCTA ATGACAGCAA AATCATCACC GTGGGTAAAA ATAACAAACC 1800
 AGACGGTACT CAAGTCAACA CCCTAAAAC TAAAGGTGAA AACGGTGTG ATGTTACAAC 1860
 CGAAACAAAT GGTACAGTTA CCTTTGGGCT TAACCAAAT AACGGTCTGA CCGTTGGCAA 1920
 CAGCACCTTA AACAACGATG GCTTATCTGT TAAAAACACC AATAGTAACA AACAAATCCA 1980
 AGTCGGTGCT GATGGCATT AATTTACTGA TATCAGCAAT AGTAAGCCAG GTGCTGGCAT 2040
 TGAAAATACC ACTCGCATT CCAGAGACGG TATTGGTTTT GCTAATAATA CTGGTTCATT 2100
 GGATGCAAAC AAACCCCGCC TAACCCCAAC TGGCATTAA GCAGGTGGTA AAGAGCTGAC 2160
 CAATGTCCAA TCTGCCATTA ACCCTGCTAC CAATGGTGGG CAGCTAGACT TTATGAACCG 2220
 CCTAAGCACT GCTAATACCG AAAAATCAGG CTCTGCCGCC ACCATTAAAG ACTTATACAA 2280
 CCTATCACAA GTACCGCTGA CCTTTGCAGG TGATACAGGT CCTAATGTCA CCAAAAAACT 2340
 GGGCGAGATT TTAAAGGTTA AAGGTGGTAA AACCACAGCT GATGATTTAA CCAAAAATAA 2400
 CATCGGTGTG GTGGCTGATA GTACCGATAA TAGCTTAACC GTTAACTTG CTAAACTTTT 2460
 AAGCGATCTT GATGCGGTTA ATACTAAAAC CCTAACTGCC AGCGATAAAG TTACCGTAGA 2520
 CAGTGGCAAC AACACCGCTA AGCTACAAAA TGGTGATTTA ACCTTTAGCA AACAAAATAC 2580
 AGGTGCTACC CCTGCCACCA ACAGCAAAAC CATCTATGGC GTTGATGGCT TGAAGTTTAC 2640
 TGATAACAAT GGTATAGCAC TTGACGGCAC AACTTACATC ACCAAAGACA AAGTTGGCTT 2700

TGCTAAGCAA GATGGTTCAC TTGATAAAAG CAAACCTTAT CTTGATAAGG ACAAGCTAAA	2760
AGTGGGTGAA GTTGAGATTA CCACCAACGG CATTAAATGCA GGTGGTAAAG CCATCACAGG	2820
ACTAAGCAAT ACCCTAACCG ATGCCACCAA CGCAACAACA GGGCATGTAA CTCAATTGGG	2880
TATCGTTGAT AGTACTGACA AAACCCGTGC CGCCAGCATT GGTGATGTGC TAAACGCAGG	2940
CTTTAACCTA AAAAATAATG GTGACGCCAA AGACTTTGTC TCCACTTATG ACACTGTTGA	3000
TTTTATCAAT GGCAATGCCA CCACCGCTAA AGTCACTTAT GATGGCAAAG CCAGTAAAGT	3060
GGCGTATGAT GTCAATGTGG ATGGTACAAC CATTCACTA ACAGGCGCTG ATGGCAATAA	3120
AAACCAAATT GGCCTAAAA CCACCACACT GACCAAAACA GATGCTAAAG GTGATAAAGC	3180
AATTAACTTT AGTGTTAACT CTGGTGATGA CAAAGCCCTT ATTAACGCCA AAGACATCGC	3240
CGACAATCTA AACACCCTAG CTGGTGAAAT TCGCAACACC AAAGGCACAG CAGACACCGC	3300
CCTACAAACC TTTCAAGTCA AAAAAGTCAA AGAAAATGGT GATGATGATA ATGACGCTGA	3360
CACCATCACC GTGGGTAAAG ATGCAAAAAC CAATCAAGTC AACACCCTAA AACTCAAAGG	3420
TAAAAACGGT CTTGATATTC AAACCAATAA AGATGGTACG GTTACCTTTG GCATTAACAC	3480
CCAAAGCGGT CTAAAGCCG GCAACAACAC CACTCTAAAC AACAATGGCT TGTCTATTAA	3540
AAACACCGCT GGTAACGAAC AAATCCAAGT CGGTGCTGAT GGCCTGAAGT TTGCCAAGGT	3600
TAATAATGGT GTTGTAGGTG CTGGCATTGA TGGCACAAC CGCATTACCA GAGATGAAAT	3660
TGGCTTTGCT GGGACTAATG GCTCACTTGA TAAAAGCAAA CCCACCTAA GCAAAGACGG	3720
CATTAACGCA GGTGGTAAAA AGATTACCAA CATTCAATCA GGTGAGATTG CCCAAAACAG	3780
CAATGATGCT GTGACAGGCG GCAAGATTGA TGATTTAAAA ACCGAACCTG AAAACAAAAT	3840
CAGCAGTACT GCCAAAACAG CACAAAACCTC ATTACACGAA TTCTCAGTAG CAGATGAACA	3900
AGGTAATAAC TTTACGGTTA GTAACCCTTA CTCCAGTTAT GACACCTCAA AGACCTCTGA	3960
TGTCATCACC TTTGCAGGTG AAAACGGCAT TACCACCAAG GTAAATAAAG GTGTGGTGCG	4020
TGTGGGCATT GACCAAACCA AAGGCTTAAC CACGCCTAAG CTGACCGTGG GTAATAATAA	4080
TGGCAAAGGC ATTGTCATTG ACAGCCAAAA TGGTCAAAAT ACCATCACAG GACTAAGCAA	4140
CACTCTAGCT AATGTTACCA ATGATAAAGG TAGCGTACGC ACCACAGAAC AGGGCAAGAT	4200
AATCAAAGAC GAAGACAAAA CCCGTGCCGC CAGCATTGTT GATGTGCTAA GCGCAGGCTT	4260
TAACTTGCAA GGCAATGGTG AAGCGTTGA CTTTGTCTCC ACTTATGACA CTGTCAACTT	4320
TGCCGATGGC AATGCCACCA CCGCTAAGGT GACCTATGAT GACACAAGCA AAACCAGTAA	4380
AGTGGTCTAT GATGTCAATG TGGATGATAC AACCATTGAA GTTAAAGATA AAAAAGTTGG	4440

CGTAAAAACC ACCACATTGA CCAGTACTGG CACAGGTGCT AATAAATTTG CCCTAAGCAA	4500
TCAAGCTACT GGCATGCGC TTGTCAAGGC CAGTGATATC GTTGCTCATC TAAACACCTT	4560
ATCTGGCGAC ATCCAACTG CCAAAGGGGC AAGCCAAGCG AACAGCTCAG CAGGCTATGT	4620
GGATGCTGAT GGCAATAAGG TCATCTATGA CAGTACCGAT AACAAGTACT ATCAAGCCAA	4680
AAATGATGGC ACAGTTGATA AAACCAAAGA AGTTGCCAAA GACAACTGG TCGCCCAAGC	4740
CCAAACCCCA GATGGCACAT TGGCTCAAAT GAATGTCAA TCAGTCATTA ACAAAGAACA	4800
AGTAAATGAT GCCAATAAAA AGCAAGGCAT CAATGAAGAC AACGCCTTTG TTAAAGGACT	4860
TGAAAAAGCC GCTTCTGATA ACAAACCAA AAACGCCGCA GTAAGTGTGG GTGATTTAAA	4920
TGCCGTTGCC CAAACACCGC TGACCTTTGC AGGGGATACA GGCACAACGG CTAAAAACT	4980
GGGCGAGACT TTGACCATCA AAGGTGGGCA AACAGACACC AATAAGCTAA CCGATAATAA	5040
CATCGGTGTG GTAGCAGGTA CTGATGGCTT CACTGTCAA CTTGCCAAAG ACCTAACCAA	5100
TCTTAACAGC GTTAATGCAG GTGGTACCAA AATTGATGAC AAAGGCGTGT CTTTTGTAGA	5160
CTCAAGCGGT CAAGCCAAAG CAAACACCCC TGTGCTAAGT GCCAATGGGC TGGACCTGGG	5220
TGGCAAGGTC ATCAGCAATG TGGGCAAAGG CACAAAAGAC ACCGACGCTG CCAATGTACA	5280
ACAGTTAAAC GAAGTACGCA ACTTGTTGGG TCTTGGAAT GCTGGTAATG ATAACGCTGA	5340
CGGCAATCAG GTAAACATTG CCGACATCAA AAAAGACCCA AATTCAGGTT CATCATCTAA	5400
CCGCACTGTC ATCAAAGCAG GCACGGTACT TGGCGGTAAA GGTAATAACG ATACCGAAAA	5460
ACTTGCCACT GGTGGTGTAC AAGTGGGCGT GGATAAGAC GGCAACGCTA ACGGCGATTT	5520
AAGCAATGTT TGGGTCAAAA CCCAAAAAGA TGGCAGCAA AAAGCCCTGC TCGCCACTTA	5580
TAACGCCGCA GGTGAGACCA ACTATTTGAC CAACAACCCC GCAGAAGCCA TTGACAGAAT	5640
AAATGAACAA GGTATCCGCT TCTTCCATGT CAACGATGGC AATCAAGAGC CTGTGGTACA	5700
AGGGCGTAAC GGCATTGACT CAAGTGCCTC AGGCAAGCAC TCAGTGGCGA TAGGTTTCCA	5760
GGCCAAGGCA GATGGTGAAG CCGCCGTTGC CATAGGCAGA CAAACCCAAG CAGGCAACCA	5820
ATCCATCGCC ATCGGTGATA ACGCACAAGC CACAGGCGAT CAATCCATCG CCATCGGTAC	5880
AGGCAATGTG GTAGCAGGTA AGCACTCTGG TGCCATCGGC GACCCAAGCA CTGTTAAGGC	5940
TGATAACAGT TACAGTGTGG GTAATAACAA CCAGTTTACC GATGCCACTC AAACCGATGT	6000
CTTTGGTGTG GGCAATAACA TCACCGTGAC CGAAAGTAAC TCGGTTGCCT TAGGTTCAAA	6060
CTCTGCCATC AGTGCAGGCA CACACGAGG CACACAAGCC AAAAAATCTG ACGGCACAGC	6120
AGGTACAACC ACCACAGCAG GTGCAACCGG TACGGTTAAA GGCTTTGCTG GACAAACGGC	6180

GGTTGGTGCG GTCTCCGTGG GTGCCTCAGG TGCTGAACGC CGTATCCAAA ATGTGGCAGC 6240
 AGGTGAGGTC AGTGCCACCA GCACCGATGC GGTCAATGGT AGCCAGTTGT ACAAAGCCAC 6300
 CCAAGGCATT GCCAACGCAA CCAATGAGCT TGACCATCGT ATCCACCAAA ACGAAAATAA 6360
 AGCCAATGCA GGGATTTTCAT CAGCGATGGC GATGGCGTCC ATGCCACAAG CCTACATTCC 6420
 TGGCAGATCC ATGGTTACCG GGGGTATTGC CACCCACAAC GGTCAAGGTG CGGTGGCAGT 6480
 GGGACTGTCTG AAGCTGTCTGG ATAATGGTCA ATGGGTATTT AAAATCAATG GTTCAGCCGA 6540
 TACCCAAGGC CATGTAGGGG CGGCAGTTGG TGCAGGTTTT CACTTTTAAG CCATAAATCG 6600
 CAAGATTTTA CTTAAAAATC AATCTCACCA TAGTTGTATA AAACAGCATC AGCATCAGTC 6660
 ATATTACTGA TGCTGATGTT TTTTATCACT TAAACCATTT TACCGCTCAA GTGATTATCT 6720
 TTCACCATGA CCAAATCGCC ATTGATCATA GGTAACCTTA TTGAGTAAAT TTTATCAATG 6780
 TAGTTGTTAG ATATGGTTAA AATTGTGCCA TTGACCAAAA AATTACCGAT TTATCCCGAA 6840
 AATTTCTGAT TATGATCACT TTTCATAAAT TTCCCAATT TGTCTTTATA AATATCCCAA 6900
 GAAATGGTAT TATTTTATTG CCATCAGCAT ATGCGACAAC TCATCGTATC ATCTTTTTAT 6960
 CATAAAAATG CAAATAGGCA TATGCATTTT TTGAATTGAA CTTACGCACT GAGAGATCCC 7020
 CTCATAATTT CCCCAAAGCG TAACCATGTG TGAATAAATT TTGAGCTAGT AGGGTTGCAG 7080
 CCACGAGTAA GTCTTCCCTT GTTATTGTGT AGCCAGAATG CCGCAAAACT TCCATGCCTA 7140
 AGCGAACTGT TGAGAGTACG TTTCGATTTC TGACTGTGTT AGCCTGGAAG TGCTTGTCCTC 7200
 AACCTTGTTT CTGAGCATGA ACGCCCGCAA GCCAACATGT TAGTTGAAGC ATCAGGGCGA 7260
 TTAGCAGCAT GATATCAAAA CGCTCTGAGC TGCTCGTTTC GCTATGGCGT AGGCCTAGTC 7320
 CGTAGGCAGG ACTTTTCAAG TCTCGGAAGG TTTCTTCAAT CTGCATTTCG TTCGAATAGA 7380
 TATTAACAAG TTGTTTGGGT GTTCGAATTT CAACAGGTAA GTTAGTTGCT AGAATCCATG 7440
 GCTCCTTTGC CGACGCTGAG TAGATTTTAG GTGACGGGTG GTGACAATGA GTCCGTGTCTG 7500
 AGCGCTGATT TTTTCGGCCT TTAGAGCGAG ATTTATACAA TAGAATTTGG CATGAGATTG 7560
 GATTGCTTTT AGTCAGCCTC TTATAGCCTA AAGTCTTTGA GTGACTAGAT GACATATCAT 7620
 GTAAGTTGCT GATAGGTTTC CAGTTTTCG CTCCTAGGTC TGCATATTGT ACTTTTCCTC 7680
 TTAAGTACTG TAACCACTAC CAACCCAGCT TCTCAACGGA TTTATACCAT GGCACCTTAA 7740
 AGCCAGCATC ACTGACAATG AGCGGTGTGG TGTTACTCGG TAGAATGCTC GCAAGGTCGG 7800
 CTAGAAATTG GTCATGAGCT TTCTTTGAAC ATTGCTCTGA AAGCGGGAAC GCTTTCTCAT 7860
 AAAGAGTAAC AGAACGACCG TGTAGTGCGA CTGAAGCTCG CAATACCATA AGCCGTTTTT 7920

GCTCACGGAT ATCAGACCAG TCAACAAGTA CAATGGGCAT CGTATTGCCC GAACAGATAA	7980
AGCTAGCATG CCAACGGTAT ACAGCGAGTC GCTCTTTGTG GAGGTGACGA TTACCTAACA	8040
ATCGGTCGAT TCGTTTGATG TTATGTTTTG TTCTCGCTTT GGTGGCAGG TTACGGCCAA	8100
GTTCGGTAAG AGTGAGAGTT TTACAGTCAA GTAAGGCGTG GCAAGCCAAC GTTAAGCTGT	8160
TGAGTCGTTT TAAGTGTAAT TCGGGGCAGA ATTGGTAAAG AGAGTCGTGT AAAATATCGA	8220
GTTCGCACAT TTTGTTGTCT GATTATTGAT TTTTGGCGAA ACCATTTGAT CATATGACAA	8280
GATGTGTATC TACCTTAAC TAATGATTTT GATAAAATC ATTAGGGGAT TCATCAGACT	8340
TACGCATCTT TCATTATGGG AATTAGGTCA GTAATTATGA CAAAAAATTA TGCATTATTA	8400
TCCGTCTCAG ATAAAACGCA AATCGTTGAA TTTGCCCAAG GTTTGGTAGA ATCTGGCTTT	8460
GGTATTTTAT CCACAGGTGG TACTTTTAAA CTCTTAAAAG AACATGGGAT TGACGCCATT	8520
GAGGTTTCTG CCCATACAGG TTTTGCTGAA ATGATGGATG GTCGTGTTAA GACCCTACAT	8580
CCCAAAATTC ATGGTGGTAT TTTGGGCCGT CGTGGCATTG ATGATGCCAT TATGAATGAA	8640
CATGGCATTG ATCGCATTGA TATCGTTGTC GTGAATTTAT ATCCATTTGC CAACACGGTC	8700
GCCAAAGACG GTGTTGTTAT GTCTGATGCG ATTGAAAATA TTGATATTGG TGGGCCTGCT	8760
ATGGTACGCT CAGCCGCCAA AAATCATGCC CATGTTGGTA TTATCACCAG CCCAATGAC	8820
TACTCACGCA TCCTAGATGA ACTAAAAAC CAAGGTCATT TAAGCCACAA CACTCGTTTT	8880
GATTTGGCAG TCAAAGCATT TGAACACACT GCCGCCTATG ATGGTATGAT TGCCAGCTGG	8940
CTAGGTGCAC GCTTACCAGT GGATAAAGAG ACGGCACCCA GTGATGATGC CACTGCAACC	9000
ACTCAATTTT CACGCACTTT TAATCACCAA TTCACCAAAG CACAAGAGCT TAGATATGGC	9060
GAAAACCCAC ATCAGTCAGC AGCCTTTTAT GTAGATGATC ATGCAACAGA AGCGTCTGTT	9120
GCGACTGCAC AGCAATTACA AGGTAAAGCG TTGTCTTATA ATAATATTGC TGATACCGAT	9180
GCGGCACTTG AGTGTGTCAA ATCTTTTACC ACGCCTGCTT GTGTGATTGT CAAACATGCC	9240
AATCCTTG TG GTTTGCAAC ATCAGAAAAC GGTATTTTAG ATGCTTATCA CTTAGCATAT	9300
GCAACCGATC CTGAATCTGC CTTTGGTGGC ATTATTGCCT TTAACCGAGA ATTAGACAGT	9360
GATACAGCCC GTACCATCGT TGAGCGTCAA TTTGTTGAAG TCATCATCGC ACCAAGCATC	9420
GCTGAAGGTG TTCTAGAGCG GCCGCGGGCC CATCGATTTT CCACCCGGGT GGGGTACCAG	9480
GTAAGTGAC CCAATTCGCC CTATAGTGAG TCGTATTACA ATTCACTGGC CGTCGTTTTA	9540
CA	9542

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2123 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asn His Ile Tyr Lys Val Ile Phe Asn Lys Ala Thr Gly Thr Phe
 1 5 10 15
 Met Ala Val Ala Glu Tyr Ala Lys Ser His Ser Thr Gly Gly Gly Ser
 20 25 30
 Cys Ala Thr Gly Gln Val Gly Ser Val Arg Thr Leu Ser Phe Ala Arg
 35 40 45
 Ile Ala Ala Leu Ala Val Leu Val Ile Gly Ala Thr Leu Asn Gly Ser
 50 55 60
 Ala Tyr Ala Gly Ile Gly Ile Ser Glu Ala Asp Gly Gly Lys Gly Gly
 65 70 75 80
 Ala Asn Ala Arg Gly Asp Lys Ser Ile Ala Ile Gly Asp Ile Ala Gln
 85 90 95
 Ala Leu Gly Ser Gln Ser Ile Ala Ile Gly Asp Asn Lys Ile Val His
 100 105 110
 Asn Ser Asn Asn Asn Ala Asn Ile Gly Ala Lys Ala Ser Gly Asn Glu
 115 120 125
 Ser Ile Ala Ile Gly Gly Asp Val Leu Ala Ser Gly His Ala Ser Ile
 130 135 140
 Ala Ile Gly Ser Asp Asp Leu Tyr Leu Lys Lys Glu Thr Val Gln Gln
 145 150 155 160
 Ile Ser Glu Leu Leu Pro Ile Ile Arg Gly Gln Lys Ala Leu Asn Asp
 165 170 175
 Ile Tyr Gln Leu Ala Asp Thr Asn Leu Gln Lys Tyr Arg Arg Thr His
 180 185 190
 Ala Gln Gly His Ala Ser Thr Ala Val Gly Ala Met Ser Tyr Ala Lys
 195 200 205
 Gly His Phe Ser Asn Ala Phe Gly Thr Arg Ala Thr Ala Glu Gly Thr
 210 215 220
 Tyr Ser Leu Ala Val Gly Leu Thr Ala Thr Ala Lys Ala Ala Ser Ser
 225 230 235 240
 Ile Ala Val Gly Ser Asn Ala Gln Ala Ile Gly Phe Ala Ala Thr Ala
 245 250 255
 Val Gly Gly Ser Thr Gln Val Asn Leu Asn Arg Gly Ile Ala Leu Gly
 260 265 270
 Phe Gly Ser Gln Val Leu Gln Lys Asp Asn Asp Val Asn Ala Ala Asn
 275 280 285
 Val Arg Ala Tyr Ala Pro Asp Asp Asn Gln Pro Ile Asp Asn Arg Tyr
 290 295 300
 Lys Ala Thr Phe Lys Asn Gly Ala Thr Asp Val Phe Ser Ile Gly Asn
 305 310 315 320
 Ser Asn Gly Asn Asp Ser Ile Arg Arg Lys Ile Ile Asn Val Gly Ala
 325 330 335
 Gly Ser Ala Asp Thr Asp Ala Val Asn Val Ala Gln Leu Lys Glu Ala
 340 345 350
 Val Arg Leu Ala Asn Arg Gln Ile Thr Phe Lys Gly Asp Asp Ser Asn
 355 360 365
 Asn Arg Val Glu Lys Gly Leu Gly Lys Thr Leu Thr Ile Thr Gly Gly
 370 375 380

Ala Gln Thr Ser Ala Leu Thr Asp His Asn Ile Gly Val Val Gln Asn
 385 390 395 400
 Gly Asp Gly Leu Lys Val Gln Leu Ala Glu Thr Leu Thr Ser Leu Lys
 405 410 415
 Met Val Thr Thr Glu Asn Leu Thr Ala Asn Glu Lys Val Thr Val Gly
 420 425 430
 Lys Thr Arg Leu Thr Thr Asp Lys Ile Gly Phe Thr Asn Asp Met Asn
 435 440 445
 Gly Ile Asp Glu Ser Lys Pro Tyr Leu Asp Lys Asp Thr Gly Ile His
 450 455 460
 Ala Gly Gly Gln Lys Ile Thr Lys Leu Thr Ala Gly Val Val Asp Asp
 465 470 475 480
 Asp Ala Ala Thr Tyr Gly Gln Leu Lys Lys Val Asn Gln Thr Ala Glu
 485 490 495
 Ser Ala Leu Gln Thr Phe Thr Val Lys Lys Val Asp Lys Asn Gly Asn
 500 505 510
 Asp Ala Asn Asp Ser Lys Ile Ile Thr Val Gly Lys Asn Asn Lys Pro
 515 520 525
 Asp Gly Thr Gln Val Asn Thr Leu Lys Leu Lys Gly Glu Asn Gly Val
 530 535 540
 Asp Val Thr Thr Glu Thr Asn Gly Thr Val Thr Phe Gly Leu Asn Gln
 545 550 555 560
 Asn Asn Gly Leu Thr Val Gly Asn Ser Thr Leu Asn Asn Asp Gly Leu
 565 570 575
 Ser Val Lys Asn Thr Asn Ser Asn Lys Gln Ile Gln Val Gly Ala Asp
 580 585 590
 Gly Ile Thr Phe Thr Asp Ile Ser Asn Ser Lys Pro Gly Ala Gly Ile
 595 600 605
 Glu Asn Thr Thr Arg Ile Thr Arg Asp Gly Ile Gly Phe Ala Asn Asn
 610 615 620
 Thr Gly Ser Leu Asp Ala Asn Lys Pro Arg Leu Thr Pro Thr Gly Ile
 625 630 635 640
 Asn Ala Gly Gly Lys Glu Leu Thr Asn Val Gln Ser Ala Ile Asn Pro
 645 650 655
 Ala Thr Asn Gly Gly Gln Leu Asp Phe Met Asn Arg Leu Ser Thr Ala
 660 665 670
 Asn Thr Glu Lys Ser Gly Ser Ala Ala Thr Ile Lys Asp Leu Tyr Asn
 675 680 685
 Leu Ser Gln Val Pro Leu Thr Phe Ala Gly Asp Thr Gly Pro Asn Val
 690 695 700
 Thr Lys Lys Leu Gly Glu Ile Leu Lys Val Lys Gly Gly Lys Thr Thr
 705 710 715 720
 Ala Asp Asp Leu Thr Lys Asn Asn Ile Gly Val Val Ala Asp Ser Thr
 725 730 735
 Asp Asn Ser Leu Thr Val Lys Leu Ala Lys Thr Leu Ser Asp Leu Asp
 740 745 750
 Ala Val Asn Thr Lys Thr Leu Thr Ala Ser Asp Lys Val Thr Val Asp
 755 760 765
 Ser Gly Asn Asn Thr Ala Lys Leu Gln Asn Gly Asp Leu Thr Phe Ser
 770 775 780
 Lys Gln Asn Thr Gly Ala Thr Pro Ala Thr Asn Ser Lys Thr Ile Tyr
 785 790 795 800
 Gly Val Asp Gly Leu Lys Phe Thr Asp Asn Asn Gly Ile Ala Leu Asp
 805 810 815
 Gly Thr Thr Tyr Ile Thr Lys Asp Lys Val Gly Phe Ala Lys Gln Asp
 820 825 830
 Gly Ser Leu Asp Lys Ser Lys Pro Tyr Leu Asp Lys Asp Lys Leu Lys
 835 840 845

Lys Asp Gly Ser Lys Lys Ala Leu Leu Ala Thr Tyr Asn Ala Ala Gly
 1780 1785 1790
 Gln Thr Asn Tyr Leu Thr Asn Asn Pro Ala Glu Ala Ile Asp Arg Ile
 1795 1800 1805
 Asn Glu Gln Gly Ile Arg Phe Phe His Val Asn Asp Gly Asn Gln Glu
 1810 1815 1820
 Pro Val Val Gln Gly Arg Asn Gly Ile Asp Ser Ser Ala Ser Gly Lys
 825 1830 1835 1840
 His Ser Val Ala Ile Gly Phe Gln Ala Lys Ala Asp Gly Glu Ala Ala
 1845 1850 1855
 Val Ala Ile Gly Arg Gln Thr Gln Ala Gly Asn Gln Ser Ile Ala Ile
 1860 1865 1870
 Gly Asp Asn Ala Gln Ala Thr Gly Asp Gln Ser Ile Ala Ile Gly Thr
 1875 1880 1885
 Gly Asn Val Val Ala Gly Lys His Ser Gly Ala Ile Gly Asp Pro Ser
 1890 1895 1900
 Thr Val Lys Ala Asp Asn Ser Tyr Ser Val Gly Asn Asn Asn Gln Phe
 905 1910 1915 1920
 Thr Asp Ala Thr Gln Thr Asp Val Phe Gly Val Gly Asn Asn Ile Thr
 1925 1930 1935
 Val Thr Glu Ser Asn Ser Val Ala Leu Gly Ser Asn Ser Ala Ile Ser
 1940 1945 1950
 Ala Gly Thr His Ala Gly Thr Gln Ala Lys Lys Ser Asp Gly Thr Ala
 1955 1960 1965
 Gly Thr Thr Thr Ala Gly Ala Thr Gly Thr Val Lys Gly Phe Ala
 1970 1975 1980
 Gly Gln Thr Ala Val Gly Ala Val Ser Val Gly Ala Ser Gly Ala Glu
 985 1990 1995 2000
 Arg Arg Ile Gln Asn Val Ala Ala Gly Glu Val Ser Ala Thr Ser Thr
 2005 2010 2015
 Asp Ala Val Asn Gly Ser Gln Leu Tyr Lys Ala Thr Gln Gly Ile Ala
 2020 2025 2030
 Asn Ala Thr Asn Glu Leu Asp His Arg Ile His Gln Asn Glu Asn Lys
 2035 2040 2045
 Ala Asn Ala Gly Ile Ser Ser Ala Met Ala Met Ala Ser Met Pro Gln
 2050 2055 2060
 Ala Tyr Ile Pro Gly Arg Ser Met Val Thr Gly Gly Ile Ala Thr His
 065 2070 2075 2080
 Asn Gly Gln Gly Ala Val Ala Val Gly Leu Ser Lys Leu Ser Asp Asn
 2085 2090 2095
 Gly Gln Trp Val Phe Lys Ile Asn Gly Ser Ala Asp Thr Gln Gly His
 2100 2105 2110
 Val Gly Ala Ala Val Gly Ala Gly Phe His Phe
 2115 2120

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Thr Val Leu Gly Gly Lys
1 5

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gly Ile Gly Ile Ser Glu Ala Asp Gly Gly Lys Gly Gly Ala Asn Ala
1 5 10 15
Arg Gly Asp Lys Ser Ile Ala Ile Gly Asp Ile Ala Gln Ala Leu Gly
20 25 30
Ser Gln Ser Ile Ala Ile Gly Asp
35 40

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Glu Ala Asp Gly Gly Lys Gly Gly Ala Asn Ala Arg Gly Asp Lys Ser
1 5 10 15
Ile Ala Ile Gly Asp Ile Ala Gln
20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly Ala Arg Gly Cys Asn Gly Ala Tyr Gly Gly Asn Gly Gly Asn Ala
1 5 10 15
Ala Arg

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Tyr Thr Gly Asn Gly Cys Asp Ala Thr Arg Thr Cys Asn Cys Cys Asp
1 5 10 15
Ala Thr

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAAGCGGACG GGGGGAAA

18

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGCGCAATG TCACCAAT

18

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 72 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAAGCGGACG GGGGGAAAGG CGGAGCCAAT GCGCGCGGTG ATAAATCCAT TGCTATTGGT
 GACATTGCGC AA

60

72

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Glu Ala Asp Gly Gly Lys Gly Gly Ala Asn Ala Arg Gly Asp Lys Ser
 1 5 10 15

Ile Ala Ile Gly Asp Ile Ala Gln
 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "oligonucleotide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCATCATTGG AAAACGTTCT TCGGGGCGAA

30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGGTCAGCTT AGGCGTGGTT

20